8-2007

Linkage Between Daily Writing Activities and Performance on Major Multiple-choice Exams

Haley Crisp Turner
University of Tennessee, Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Haley Crisp Turner entitled "Linkage Between Daily Writing Activities and Performance on Major Multiple-choice Exams." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

Robert L. Williams, Major Professor

We have read this dissertation and recommend its acceptance:

Sherry K. Bain, John W. Formby

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Haley Crisp Turner entitled "Linkage Between Daily Writing Activities and Performance on Major Multiple-choice Exams." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

Robert L. Williams, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Vice Chancellor and Dean of
Graduate Studies

LINKAGE BETWEEN DAILY WRITING ACTIVITIES AND PERFORMANCE ON MAJOR MULTIPLE-CHOICE EXAMS

A Dissertation
Presented for the
Doctor of Philosophy Degree
The University of Tennessee, Knoxville

Haley Crisp Turner
August 2007

DEDICATION

This dissertation is dedicated to my husband, Chad, my parents, Bill and Brenda Crisp, my sister Abbey, and the rest of my wonderful family, for always believing in me and inspiring me to reach higher than I ever imagined I could! I thank each of you for your unending support and encouragement.

ACKNOWLEDGEMENTS

I want to thank all of the students and professors who helped me in completing my Doctor of Philosophy Degree in School Psychology. I would like to especially thank Dr. Williams for the opportunity to teach and more importantly for being a wonderful mentor. I would also like to thank Drs. McCallum, Skinner, Bain, and Lounsbury for their support and encouragement. Finally, I want to thank all of the wonderful students that I met along the way.

Abstract

Presumably, one approach for improving exam performance in college courses is to increase the quantity and quality of pre-class preparation. Over a two-semester period, students in six sections of a large undergraduate course in human development at a large southeastern university participated in a study designed to assess the efficacy of an intervention to enhance student preparation for class. The treatment involved administration of daily five-minute writing activities based on pairs of concepts embedded in the instructor notes of each unit in the course. The principal dependent variable across semesters was student performance on unit-exam items related to content in the instructor notes and power-point slides. Student performance on these exam items was contrasted across semesters (one semester with the writing activities and one semester without the writing activities) and across written performance levels for the semester in which the writing activities were used.

In general, performance on the writing activities was a better predictor of performance on exam items related to the instructor notes and power-point slides than performance on items related strictly to issues in the reading materials. Specifically, there were four major findings: (1) exam performance on the targeted items was higher the semester when the writing activities were used than when they were not used; (2) the relationship between performance on the writing activities and exam performance was stronger for exam items related to the instructor notes/PowerPoint slides than items related only to issues in the reading materials; (3) daily writing activities were a stronger predictor of exam performance than other previously predictive course variables (pre-

course critical thinking, pre-vocabulary, and attendance); and (4) students who scored high on the writing activities scored higher on the targeted exam items than those who scored low on the writing activities.

Table of Contents

List of Tables

Table Page

List of Figures

Figure Page

Chapter I

Introduction

College instructors are continually searching for techniques to improve student performance. Increasing exam scores is paramount to improving student performance in a variety of courses. Presumably, one approach for improving exam performance is to increase the quantity and quality of pre-class preparation. Although instructors often expect students to come to class prepared for discussion, students in the Educational Psychology 210 course targeted in this study have rarely met this expectation. Even with prior access to instructor notes, students infrequently take advantage of this opportunity to become well versed in course content prior to coming to class. The instructional technique designed for this study is intended to increase students' pre-class preparation for class discussion. Presumably, by increasing pre-class preparation, students can participate in a more informed fashion in class discussion and derive more benefit from class discussion.

In principle, students' engagement in out-of-class preparation activities greatly affects in-class performance. Unfortunately, a survey conducted by the Higher Education Research Institute in 2001with 281,064 incoming college freshmen at 421 institutions found that students reported doing less studying and homework than in a similar study in 1987. In 2001, 34.9% of students reported studying or doing homework for six or more hours per week as high school seniors compared to 47% in 1987 (www.gseis.ucla.edu/heri/norms_pr_01.html). Once in college, although students spend more time on academic preparation than in high school, professors indicate that their

preparation is inadequate. A survey conducted by the Center for Postsecondary Research in 2004 with 620,000 college freshman and seniors at 850 academic institutions and 20,000 faculty members at 132 institutions found that students report spending, on average, between 9 and 14 hours a week preparing for college classes. However, faculty members believe students should study considerably more. In fact, faculty members indicated that 25 hours of preparation a week is necessary for success in college. However, only 11% of the full-time students surveyed reported spending 25 or more hours a week preparing for classes. Instructors expect students to spend about 6 hours a week preparing for each class, whereas students report spending only about 3 hours a week per class (www.iub.edu/~nsse). Given this disparity, it seems imperative that college instructors find ways to increase the amount of time students spend preparing for class.

Research has validated the efficacy of several instructional techniques in increasing pre-class preparation. Moreover, many of these techniques appear to have increased subsequent exam scores. One line of research has evaluated the comparative effects of random versus voluntary oral questioning on class preparation (McDougall & Granby, 1996). Results showed that students who expected the instructor to randomly call on them during class to answer questions over the assigned reading material completed more reading before class, recalled more information from the reading, and indicated having more confidence in their understanding of the reading material than peers in a class that employed voluntary responding to instructor questions. This finding supports the instructional technique used in the course targeted in this study. In this course, after

instructors pose a question, they typically call on students to respond rather than waiting on them to volunteer an answer.

Similarly, researchers have considered the impact of students' writing daily questions over assigned readings on student performance. For example, Shaw (2001) had students engage in a "make-the-quiz" exercise whereby students were responsible for writing five multiple-choice questions for each assigned reading. Students earned course credit for a particular question only if it was selected by the instructor to be included on a weekly quiz. Anecdotal evidence from students indicated that they read the material more thoroughly and displayed greater understanding of course concepts in class under the "make-the-quiz" condition than under typical classroom procedures. The instructor further noted that students displayed more certainty in discussing assigned material in class under the "make-the-quiz" condition.

Kerkman, Kellison, Pinon, Schmidt, and Lewis (1994) empirically demonstrated the effectiveness of question writing on quiz performance. Students in the experimental group were told to construct and submit one multiple-choice question for each assigned reading. The written questions contributed 20% toward the course grade. Students enrolled in another section of the same course with the same instructor served as the control condition. Both groups were given multiple-choice pop quizzes composed of questions constructed by students in the experimental group. No measures were taken to ensure that quiz questions were not shared between students in the two sections prior to the in-class quizzes. Results showed that students in the experimental group had significantly higher quiz scores than students in the control group.

Tai-Seale and Thompson (2000) have evaluated the effects of assigned conversations on pre-class preparation. In this method, students are assigned a specific portion of the assigned reading material to discuss in class. Among students in the assigned-conversation class, 91% either strongly agreed or agreed with the statement that the assigned conversations encouraged them to prepare and read before coming to class. When compared to another class section without assigned conversations, students in the experimental class were more likely to indicate that they prepared before coming to class, learned a lot in the class, and actively participated in class activities and discussions. About 65% of students in the assigned conversation condition indicated a preference for assigned conversations compared to traditional lectures.

Another extensively researched instructional technique for improving class preparation is "Just-in-Time Teaching" (JiTT) (Simkins & Maier, 2004; www.jitt.org). JiTT involves student completion of out-of-class exercises over information to be covered the next class. The JiTT exercises are short assignments completed and submitted electronically. Student answers are typically no longer than a paragraph, and instructors evaluate the responses before class. In this way, instructors can adjust the upcoming class discussion based on students' understanding of the concepts included in the JiTT exercise. JiTT exercises are thought to encourage pre-class preparation by focusing student attention on the most critical course concepts. Results of a study comparing exam performance of students with and without JiTT activities showed that students performed better on JiTT-related exam items when they engaged in the out-of-class JiTT exercises. Qualitative data suggested that most students regarded the JiTT exercises as more influential in their decision to read the textbook than was studying for

tests or elaborating on class lectures.

Researchers have also evaluated the efficacy of quizzes on student pre-class preparation and subsequent course performance. To encourage pre-class preparation, Thorne (2000) randomly administered extra-credit pop quizzes. Each pop quiz could potentially earn students up to one-point extra credit. Qualitative evidence from course evaluations suggested that students liked the extra-credit pop quizzes. Further, Thorne reported that the pop quizzes encouraged pre-class preparation and gave students a model for exam items. On course evaluations, 25% of students indicated that they liked the extra-credit pop quizzes compared to other course features.

Tuckman (1996) has provided additional findings regarding the differential effects of weekly quizzes and homework assignments on student performance. In the first experiment, classes were randomly assigned to one of three treatment conditions. In one class, students engaged in weekly 15-minute fill-in-the-blank quizzes over the assigned textbook chapter. In the second class, students completed weekly homework assignments requiring them to identify, define, and provide one-sentence elaborations for the 21 most important terms in the assigned chapter. Under both conditions, performance on either the weekly quiz or homework assignment constituted one-fourth of the final grade. The third class, which served as the control condition, involved no quizzes or homework assignments. To assess student performance under the three conditions, the instructor gave identical 50-item multiple-choice exams. The exams involved a higher level of cognitive understanding than the quizzes and homework assignments, which focused primarily on factual recall. Results showed that exam performance was significantly better for students in the weekly quiz condition than students in the other two conditions.

Mean exam performance for students in the weekly quiz condition was 82.8%, compared to 71.6% for students in the homework condition and 66.9% for students in the control condition.

A second experiment was conducted to determine whether student's GPA predicted performance under the two treatment conditions. Like the first study, results of this study found that students in the weekly quiz condition outperformed students in the homework condition. However, a significant interaction was found between GPA and treatment condition. Specifically, students with low GPAs performed significantly better on exams under the weekly quiz condition than students in the homework condition. Tuckman (1996) further investigated whether study time differed across the two conditions. Surprisingly, he found that students in the homework condition reported spending 38 minutes more per week preparing for class than students in the quiz condition.

Ruscio (2001) has empirically demonstrated the effectiveness of random pop quizzes on pre-class preparation. At the beginning of each class, a coin was flipped to determine whether a pop quiz would be given. The quizzes were composed of one or two instructor-constructed short-answer questions based on the assigned reading. Points from the pop quizzes constituted 15% of the course grade. Results indicated that students earned full credit on 74% of the pop quizzes, and 85% of the students reported completing at least half of the assigned readings. Further, quiz grades correlated highly with the remaining 85% of the course grade. These findings suggest that random pop quizzes may increase the consistency of pre-class reading.

Conner-Greene's (2000) research provides further evidence for the efficacy of

daily essay quizzes in promoting pre-class preparation. In the study, the experimental class was assessed with daily essay quizzes. The quizzes consisted of one or two questions over the assigned reading material and information presented in the previous class session. The control class had four scheduled tests over the course of the semester. The author reported that although the experimental and control conditions involved different courses, the amount of assigned reading in each class was comparable. Results showed that 92% of students in the experimental essay quiz group reported completing the reading by the scheduled date, compared to only 12% of students in the control condition. Seventy-two percent of students in the control class reported rarely or never reading the assignments by their scheduled date, compared to zero percent of students in the daily essay-quiz group. Students in both groups generally thought that daily essay quizzes would require more pre-class preparation time, more reading, and a better understanding of the material than having four scheduled exams. They further believed that daily essay quizzes would improve class participation. Ninety-six percent of students in the daily quiz condition indicated that they preferred the daily quiz schedule to four scheduled exams. The daily essay quizzes seemed to reduce procrastination, given that more students completed the reading by the assigned date.

Other researchers have investigated the comparative effects of immediate and delayed quizzes on procrastination, on-task in-class behavior, and student performance (DeRoma et al., 2003). Procrastination was measured by Tuckman's (1998) 35-item scale. Procrastination was measured in three areas: avoiding activities, delaying activities, and attributing bad results to something other than oneself. In this study, students were administered quizzes over a lecture either at the end of a class period

(immediate) or at the beginning of the class period following a lecture (delayed). Using Momentary Time Sampling, no significant between-groups difference was found for student on-task behavior. Thus, immediate quizzes were not superior to delayed quizzes in increasing students' attention during class. A significant difference, however, did emerge on quiz performance under the two conditions. Quiz scores were significantly higher under the immediate condition than the delayed condition. Additionally, students' Procrastination Scale scores were negatively correlated with performance on the delayed quizzes. Thus, higher procrastination scores predicted lower scores on the delayed quizzes.

Tuckman (1998) has expanded the research on decreasing procrastination and increasing student performance. Students in two sections of an undergraduate course served as participants. Students in an incentive approach were given biweekly quizzes covering the assigned chapters, whereas students in the strategy approach were instructed to outline the assigned chapters. Students in both conditions could earn an equivalent amount of course credit for their performance on the quizzes or outlines. Students in both conditions were given an identical 65-item multiple-choice final exam to assess performance. Results showed that students in the biweekly quiz condition performed significantly higher on the final exam than students in the outline condition. There was also a significant interaction between procrastination scores and condition. That is, students in the biweekly quiz condition who scored high on the procrastination scale did significantly better on the final exam than high procrastinators in the outline condition. Results of this study demonstrate the superiority of essay quizzes over an outlining procedure for promoting exam performance, especially for high procrastinators.

Another line of research has evaluated the link between exam frequency and student performance. Research has generally shown that students perform better when tested more frequently. Fulkerson and Martin (1981) investigated the effects of exam frequency on various student outcomes. Results showed that with equivalent exam items, students who took 25-item tests every two weeks significantly outperformed students who took 50-item tests every four weeks. Anecdotal evidence provided further support for more frequent testing. Students who took more frequent exams rated the instructor better on every question on a course-evaluation instrument.

Although most research seems to support the efficacy of pre-class preparation on subsequent exam performance, some research indicates that this preparation may be ineffectual in improving exam performance. Spies and Wilkin (2004) evaluated the effects of pre-class preparation of legal cases by pharmacy students. They hypothesized that the preparation would lead to better understanding and discussion of the cases in class. At the beginning of the course, students were divided into four groups. Prior to class, one group was given a legal case to be discussed by all students during the next class period. During the subsequent class, the group with prior access to the case was expected to explain its facts and law implications to the rest of the class. Students in each of the groups were given several different legal cases to present over the course of the semester. However, case assignment was mutually exclusive. Essay questions related to the assigned legal cases discussed in class were included on the exams to assess student performance. Results showed that in only one case did members of the presenting group outperform their peers on the essay questions corresponding to the cases they prepared and presented. These results presumably demonstrated that pre-class preparation of legal

cases did not have a significant impact on exam performance.

Research has shown that a fundamental ingredient to increasing student's preparation for class is the promotion of deep learning (www.iub.edu~nsse). Deep learning activities can be characterized by higher-order learning, integrative learning, and reflective learning. Results show that when students report engaging in deep-learning tasks, they also report spending more time preparing for class and more time reading material out of class. The instructional technique used in the current study presumably involves a high level of deep learning. In the daily writing activities, students were expected to expand on information from their instructor notes to compare and contrast pairs of concepts. In this way, students' ability to synthesize and integrate information from the instructor notes was assessed in the writing activity. I believe the activity involved student engagement in two deep learning components: higher-order learning and integrative learning. The promotion of deep-learning in the daily writing activities logically should facilitate the use of a deep approach in preparing for the course exams. Because exam items require higher-order reasoning, use of a deep approach to learning the course material would appear to promote better exam performance.

The type of assessment techniques used in a course can affect the depth and extent of student learning. Crooks (1988) review of the research on classroom evaluation practices concluded that using assessment techniques that center around higher-order learning increases student learning and retention. Differential approaches to studying result from students' perception of the assessment task. Using Marton and Saljo's (1976) approaches to learning, students can either take a deep or surface approach to learning.

Deep approaches are those that involve making connections between concepts, inferring meaning, identifying overarching principles, and actively applying learned material. Alternately, surface approaches involve disconnected rote memorization and regurgitation. Based on his review of numerous studies showing the effects of assessment on students' choice of learning approach, Crooks postulated that deep learning should be an overriding goal for educators.

Dickie (2003) specifically found a relationship between students' learning approach and the cognitive demands of assessment. Using the Study Process Questionnaire to ascertain student's learning approach, he found that 41% of entering college students relied on surface-level or surface-achieving approaches to learning, compared to 5.6% who reported using a deep approach. The surface-level approach primarily includes blind adoption of superficial memorization. In the surface-achieving approach, the student adopts either a deep or surface-level approach depending on how they perceive the upcoming assessment. For example, if a student believes an upcoming exam will primarily require factual recall, he or she will likely adopt a surface-level approach to studying. In this study, after evaluating the assessment technique to be used, all students using the surface-achieving approach chose a surface-level approach. The deep approach involves more complex learning intentions. Results of Dickie's study indicated that when assessments were believed to involve a high level of intellectual demand, students were more likely to adopt a deep approach to learning the course material. Additionally, a negative relationship was found between performance on exams involving high intellectual demand and use of the surface approach. Conversely,

performance on an exam with low intellectual demand was negatively correlated with a deep-learning approach. That is, students were less likely to use a deep learning approach to study for a test presumed to involve a low level of intellectual demand. Results of this study support the use of assessment techniques involving a substantial amount of intellectual demand.

Research on the unit exams targeted in the current study has indicated that the items on the exams involve considerable intellectual demand. Wallace and Williams (2003) classified exam items as direct recall, comprehension, or mixed. Twenty-six percent were determined to be direct recall, involving mastery of facts that closely paralleled lecture or reading materials. Fifty-eight percent were classified as comprehension items addressing factual information presented differently on the exams than in lecture or course materials. The remaining 16% of items were classified as mixed, involving direct recall and comprehension. Student performance on the various types of exam items was correlated with performance on a generic critical thinking measure. Specifically, students' performance on the California Critical Thinking Skills Test (Facione & Facione, 1994) was correlated with total exam performance and performance on direct recall, comprehension, and mixed exam items. Findings showed scores on the critical thinking measure significantly predicted total exam performance. More specifically, comprehension exam items correlated .42 with critical thinking.

Ostensibly, college instructors are seeking efficient and effective methods for improving student preparation for class discussion. Based on the results of previous research, techniques that increase the amount of student out-of-class preparation on a

daily basis may provide other related benefits for students (such as improvement in exam performance). However, much of the previous research on out-of-class preparation is of a self-report nature, with students rating the impact of procedures designed to improve class preparation. Past research evaluating the effects of these treatment procedures has provided negligible evidence of inter-rater reliability for the evaluation procedures. Also, daily performance measures (e.g., exam questions, quizzes), as used in several previous studies, can be very labor-intensive with respect to instructor evaluations. The daily procedures may prove very effective in improving student performance but not very efficient with respect to the time required for instructors to manage these daily procedures, especially in courses with large enrollment.

In the current study, a daily instructional technique related to pre-class preparation was designed to improve performance on course exams. Given the subjective nature of much past research on techniques designed to improve pre-class preparation and subsequent exam performance, the current study involved written products that would serve as empirical evidence for pre-class preparation; plus, the study also established acceptable inter-rater reliability for the evaluation of those products. The latter was intended to ensure the replicability of the findings by other researchers.

The treatment strategy involved the administration of short daily writing activities based on critical concepts previously announced for discussion each class session. Although students constructed their written products at the beginning of most class sessions, only a few of these products were randomly selected for grading. The targeted daily writing activities were designed to encourage students to think critically about issues embedded in the course. Moreover, students' adoption of more complex learning

approaches in preparing for the daily writing activities was expected to facilitate the use of such approaches on the unit exams. Thus, the study attempted to determine if a procedure involving daily preparation for an in-class writing task could meet the criteria for both effectiveness (improve class discussion and exam performance) and efficiency (not require an inordinate amount of instructor time).

Providing credit for students' engagement in daily writing activities, of course, did not assure that they would actually engage in such activities, especially at a high level. A principal research question of the current study was whether frequent and/or high-quality engagement in the conceptually based written activities would help students prepare for major exams different in format (multiple-choice) from the writing activities. The predictive potential of the daily writing activities was compared to that of other known predictors of exam performance in the target course (e.g., critical thinking, generic vocabulary, and attendance). Prior research in the target course provides evidence for the predictive potential of pre-course critical thinking, attendance, and pre-vocabulary. Williams and Worth (2002) found that performance on a pre-course critical thinking assessment correlated .46 with exam performance. Similarly, a strong relationship has been established between exam performance and attendance. Williams and Worth found a .49 correlation between class attendance and performance on unit exams. Researchers have also found a strong correlation between exam performance and performance on a pre-vocabulary assessment designed to assess pre-course generic vocabulary, $r = .45$ (Turner & Williams, 2004).

Specifically, the current study addressed the following questions regarding the effectiveness and efficiency of the daily writing activities: (1) To what extent did student exam performance differ across the treatment (writing activities) and non-treatment (no writing activities) semesters? (2) How well did the *frequency* versus the *quality* of students' engagement in the daily writing activities predict their performance on targeted (instructor notes and power-point slides) versus non-targeted (text and journal articles) multiple-choice exam items? (3) How well did student performance on the writing activities, compared to other established exam-performance predictors, predict performance on the exam items? (4) How well did students' performance on *each day's* writing activities across units predict their *composite writing performance* in the course? (5) How well did students' performance on *each day's* writing activities across units, compared to their composite performance on writing activities in the course, predict their performance on targeted and non-targeted *exam items*? (6) To what extent did student performance on the daily writing activities vary *across days* and *across units* in the course? (7) How well did high-performing students on the daily writing activities, compared to students performing poorly on these activities, do on the targeted exam items?

Chapter II

Method

Participants

Students ($N = 299$) in six large sections of an undergraduate course in human development participated in various aspects of the study. The course is required for students entering the Teacher Preparation program at a large southeastern university. The data were collected over two semesters in classes ranging from 49 to 55 students. The sample included more women than men (about a 2 to 1 ratio), with a majority of the sample being sophomores and juniors. The self-reported grade point averages of the participants ranged from 1.8 to 4.0, with the total-sample average being 3.14.

The participants in the study were initially divided into two samples based on which semester they were enrolled in the course. In the fall 2004 semester of the course, students ($n = 153$) did not engage in daily writing activities (no treatment). Conversely, in the spring 2005 semester, students ($n = 146$) completed daily writing activities (treatment) that contributed up to 25 points towards their overall course grade. Performance on the daily writing activities potentially contributed approximately 4% to the final grade in the course. To determine if student performance on the writing activities was related to exam scores on items tied to the instructor notes, I contrasted the exam performance of students ($n = 33$) who scored in the top quartile (high-performers) on the written activities with those ($n = 30$) in the bottom quartile (low-performers) in the spring 2005 semester.

Course Structure and Treatment Strategy

The course is a gateway course for the teacher-education program. The course targets five developmental themes, with each representing a unit in the course: physical development, cognitive development, psychological development, social development, and character development. During the treatment semester, seven class periods were devoted to each unit. The treatment strategy involved student completion of writing activities on the first five days of each unit. In each unit, students could earn up to five points course credit for their performance on one of the daily writing activities. Although students engaged in the writing activity on five out of seven classes during each unit, only one daily writing activity was selected to count towards their grade in the course. On the 5th day of each unit, after the writing activity, a student in each class drew a number from 1 to 5. The writing activity to be graded was identified by the number drawn. For example, if a 2 was drawn, student's performance on the writing activity from day 2 in the unit would be evaluated for credit.

Student performance on the daily writing activities was evaluated by advanced school psychology doctoral students. The daily writing activities were taken from important concepts embedded in instructor notes made available to students. The purpose of the writing activities was to challenge students to compare and contrast important concepts in the notes by synthesizing specific information in the instructor notes. This process approximated what students were asked to do in responding to multiple-choice exam items in the course. Explicit criteria were developed for rating the students' written responses, with inter-observer agreement computed for 40% of the students' responses.

The grading criteria required at least one statement pinpointing what the two concepts had in common or how they were linked and then at least four distinct statements contrasting the concepts. Prototypic answers developed by the course supervisor aided raters in determining what to look for in student responses to the conceptual pairs. (See Appendix A for samples of these prototypic answers). Across units, the overall inter-rater agreement ranged from .86 to .91. Although only one daily writing activity per unit contributed to students' course grade, all writing activities for all units were rated for research purposes.

In the treatment semester of the course, each student was expected to purchase a study guide containing a list of conceptual pairs and instructor notes for each unit. The instructor notes provided students with a framework for class discussion. Before each class period, instructors posted several conceptual pairs listed in the study guide at the course web site. The conceptual pairs represented important issues to be discussed in class. (Conceptual pairs for all units are provided in Appendix B). Students were instructed to study the instructor notes, the on-line power-point slides linked to the notes, and the posted conceptual pairs before coming to class. Further, they were told to be able to identify what each pair had in common and at least four ways in which the concepts were different. Students were expected to use their instructor notes and power-point slides to prepare for the writing activities. Students were advised to construct a written response for each of the posted conceptual pairs prior to coming to class. Instructors expected that this pre-class practice would help students construct their written responses in class, when they were not allowed to use their instructor notes or power-point slides.

At the beginning of class, instructors chose one of the previously posted conceptual pairs for students to compare and contrast in writing.

Each unit in the course culminated with a 50-item multiple-choice exam. Performance on exam items related to the instructor notes and power-point slides was the principal dependent variable in the study. Each exam item was classified as either a notes or readings item depending on the source for the item. The instructional supervisor in the course determined where information for each exam item came from in the course materials. Items taken from the instructor notes or power-point slides were classified as notes items. Alternately, items derived from the out of class reading assignments were classified as readings items. Because the experimental daily writing activities were based on the instructor notes, the notes items on the exams were considered the targeted items and the readings items the non-targeted items. Items derived from multiple sources (i.e., notes and readings) were not included in the analysis.

Additionally, information on the internal consistency of the unit exams was computed. Cronbach Alpha's for the unit exams were as follows: .769 for A, .810 for B, .825 for C, .693 for D, and .799 for E. Reliability information was also computed for the two item types: notes and readings. Cronbach Alpha's for the notes items ranged from .536 on Unit A notes items to .750 for Unit C notes items. For the non-targeted reading items, Cronbach Alpha's ranged from .362 for Unit A reading items to .674 for Unit B reading items. Additionally, reliability statistics across all unit exams were calculated. Composite Cronbach Alpha's were .870 for the Unit Exam Total, .781 for the Notes Exam Total, and .726 for the Readings Exam Total.

Because the number of notes and readings exam items differed across unit exams, students' notes and readings exam scores for each unit, as well as their total notes and readings exam scores, were converted into *z*-scores ($M = 0$, $SD = 1$) across semesters, permitting both cross-unit and cross-semester comparisons. If scores had been normalized within-semesters, exam means across semesters would have been identical and cross-semester comparisons would have been meaningless. Thus, student exam scores from fall and spring were combined and then converted into *z*-scores.

As a point of contrast for performance on exam items pertaining to the instructor notes and power-point slides, performance on exam items related to the textual reading material also was assessed. However, in-class discussion of issues in the reading materials differed across the two semesters. In the fall 04 section of the course, a short (30-minute) review was devoted to discussion of the outside reading materials. During this review, students were permitted to ask for clarification of confusing issues in the readings. Conversely, in the spring 05 semester, no class time was spent discussing the assigned readings. Rather, students were instructed to ask their questions about reading materials through e-mail exchanges or phone messages with their instructor. However, anecdotal evidence from instructors indicates that students rarely took advantage of this opportunity. Also, on the day before each unit exam in spring 05, students were given a list of the specific study questions over the reading materials that would be addressed on the exams. This pre-exam list of readings questions was not provided for the Fall 04 students. Although performance on the readings items was included only as point of contrast for performance on the notes items, it was expected that fall 04 students would

do better on the readings items than the spring 05 students because of the in-class discussions of the readings in fall 04.

Writing-Activity Dimensions

Several dimensions of the writing activities were explored in data analyses: number, total, unit average, and daily average. The number of writing activities was simply the total number (out of 25) of writing activities a student engaged in (potentially 5 per unit) during the course. The total of the writing activities was computed by adding a student's total score earned (potentially 125) on all writing activities attempted (potentially 25 total per unit). The unit writing average was figured by totaling a student's credit earned on all writing activities within a unit and dividing that total by the number of writing activities attempted in the corresponding unit (e.g., summing all Unit A writing scores for a student and then dividing by the number of Unit A writing activities attempted). A total of the unit means for individual writing-averages also was computed for the combined units, representing the sum of a student's unit averages.

The daily writing-average score was figured by adding a student's scores on all writing activities on a particular day across units and dividing that total by the number of attempted writing activities on the corresponding day (e.g., summing all Day 1 writing-activity scores across the five units and dividing by the number of Day 1 writing activities submitted). For example, if a student did not have a writing score for Day 1 of Unit A, the first day in Unit A was not included in the computation of the student's average for Day 1 across units. On the other hand, if a student submitted 3 out of 5 writing products for Day 1 across units and earned a total of 15 points for those three

products, that student's Day 1 average would be 5 (15 divided by 3) instead of 3 (15 divided by 5). Similar to the total of average-score means across units, a total daily-writing average was computed across days (1 to 5), representing the sum of a student's daily averages.

Comparison Predictors

The predictive potential of the aforementioned writing-activity dimensions was compared with three other variables previously shown to predict performance on course exams: attendance, critical thinking, and generic vocabulary. The writing-activity dimension (i.e., number, total, or average) that proved to be the best predictor of exam performance was first determined and then contrasted with the other known predictors of exam performance in the course. The computation of the writing-activity dimensions has been described earlier in this section. In addition, attendance was checked each of the five days per unit that students did writing activities by having them sign an attendance form on the line beside their typed name. Critical thinking was assessed at the beginning of the course by administering the *Watson-Glaser Critical Thinking Appraisal-Form S* (WGCTA), perhaps the most widely used generic measure of critical thinking (Watson & Glaser, 1994). The test manual indicates that both internal consistency and test-retest reliability for Form S is .81. Students' mastery of generic vocabulary was assessed by the researchers' first identifying potentially unfamiliar words on the course exams (unit exams, practice exams, and final exam) and then asking students to choose a parallel term for each of these vocabulary words on a 50-item multiple-choice vocabulary test administered at the beginning of the course. Approximately 50% of the terms on the

vocabulary test were embedded in unit-exam items (Turner & Williams, 2004). Students were informed at least one to two days before each unit exam which vocabulary terms would appear on that exam.

Statistical Analyses

Student exam performance on the targeted notes items was expected to be better for students enrolled in the spring 2005 semester of the course (when the writing activities were used) than in fall 2004. However, because no pre-course assessment of exam performance was done the two semesters targeted in this study, I first did a retrospective analysis of student exam performance across several previous fall and spring semesters to determine equivalency of exam performance across semesters similar to those targeted in the current study. In addition, because critical thinking has proven to be one of the most powerful predictors of exam performance in previous semesters, an independent samples *t*-test was performed on student's pre-critical thinking averages for the two semesters included in the current study. Finally, to adjust for unequal numbers of notes and readings items across unit exams in the targeted semesters, student exam scores on both dimensions were converted into normalized *z*-scores across semesters. Then a repeated measures analysis of variance was used to compare exam performance on items related to the instructor notes and the readings across semesters (fall 2004 and spring 2005).

A prediction and discrimination model was used in addressing how well performance on the writing activities predicted exam performance in the spring 05 semester. Through correlation and multiple-regression analyses, I determined (a) what

dimensions of the daily writing activities (number, total, and average) best predicted performance on the targeted multiple-choice exam items; (b) how well the most predictive writing-activity dimension, compared to more conventional predictors (attendance, critical thinking, and student vocabulary), predicted performance on the targeted and non-targeted exam items; (c) how well performance on each day's writing activity across all units predicted composite writing performance on all days across units; and (d) how well each of the five days across units, compared to all 25 writing days, predicted performance on the targeted and non-targeted exam items. The last two analyses indicated how much predictive potential would be lost (a) in the writing assessment by scoring only one writing activity per unit instead of each written activity in a unit and (b) in the prediction of the targeted and non-targeted exam scores by including scores from one daily writing activity versus multiple days per unit.

Finally, an independent samples *t*-test was used to determine exam-performance differences for students who scored in the top and bottom quartiles on the most-predictive writing-activity dimension. This analysis was done to determine the apparent impact on exam performance of students' doing extremely well versus extremely poorly on the writing activities. These differences were first determined for each of the five units in the course and then for the course as a whole. Because the number of target and non-target exam items differed across units, each unit's data were converted to normalized *z*-scores across semesters. In this series of comparisons, performance-level on the writing activities served as the independent variable and targeted and non-targeted exam items as the dependent variables.

Chapter III

Results

This section presents findings in the following sequence: (a) unit-exam totals across past semesters of the course, (b) WGCTA scores for Fall 2004 and Spring 2005, (c) unit-exam scores related to targeted and non-targeted items for Fall 2004 and Spring 2005, (d) relationships of frequency versus quality of students' engagement in the daily writing activities to targeted and non-targeted multiple-choice exam items, (e) differential predictability of course variables (writing-activity dimensions, pre-course critical thinking, pre-vocabulary, and attendance) to unit exams, (f) relationship of unit writing-average scores to exam performance, (g) relationship of unit writing-average scores to total writing-average scores across units, (h) relationship of daily writing-average scores to total daily writing-average scores, (i) relationship of daily writing- average scores to exam performance, (j) differences in student performance on writing-average scores across both days and units in the course, and (k) differences in exam scores for high and low performers on writing-averages.

Student Performance across Past Semesters

Because the current study compared student performance across semesters, it was necessary to first examine the possibility of initial differences in performance across the two semesters. In the absence of pre-course test scores, other comparisons were used to demonstrate probable equivalency. For example, Table 1 shows past exam performance to be virtually equivalent across previous fall and spring semesters in the target course. Thus, there was little reason to expect the pattern of past-semester exam scores to be

Table 1

Unit Exam-Total Means across Semesters

Semester	*N*	Mean	Standard deviation	Percentage correct	Exam items
Fall 00	176	158.93	19.38	79%	40
Spring 01	145	158.26	18.24	79%	40
Fall 02	210	189.26	27.47	76%	50
Spring 03	134	190.81	26.38	76%	50
Fall 03	191	190.84	25.53	76%	50
Spring 04	202	190.09	26.56	76%	50

Note. Data were not available for Fall 01 and Spring 02. Exam score totals in 2000-01 could range from 0 to 200, whereas exam totals for 2002-04 could range from 0 to 250.

different for the two semesters targeted in the current study, had the daily-writing intervention not been employed in spring 05.

To further establish the equivalency of the two target groups (fall 04 and spring 05), student performance on the WGCTA (pre-course critical thinking test) was compared over the two semesters. As previously noted, pre-course critical thinking has proven to be among the most powerful predictors of exam performance in past semesters. The mean WGCTA score in spring 05 was 26.77, compared to 25.69 in fall 04. A *t*-test comparison failed to yield a significant difference between the two means, t (297) = 1.774, p = .08. Thus, I assumed that students had similar thinking potential across the two semesters used in the current study.

Normalized Exam Scores across Fall 04 and Spring 05 Semesters

Table 2 provides a summary of student performance on notes and readings exam items for spring 05 and fall 04. Across all five unit exams, students in the daily-writing condition (spring 05) performed better on the targeted notes exam items than students in the control condition (fall 04). Alternately, students in the daily-writing condition performed worse on the non-targeted (readings) exam items than students in the control condition. As hypothesized, exam totals across units showed that students in spring 05 did significantly better on the targeted (notes) exam items than students in fall 04 but worse on the non-targeted exam (readings) items. Although not all unit comparisons proved significantly different, the pattern of differences was entirely consistent with the differences on the exam totals. A repeated measures analysis of variance was used with total notes versus total readings scores as the repeated measure and semester as the

Table 2

Normalized Exam Means by Unit across Fall 2004 and Spring 2005 Semesters

			Notes		Readings	
	Semester	*n*	*M*	*SD*	*M*	*SD*
Exam A	Spring 05	152	.164*	.931	-.218	.939
	Fall 04	158	-.158	1.041	.209*	1.014
Exam B	Spring 05	147	.101	.936	-.134	.987
	Fall 04	152	-.098	1.052	.131*	.998
Exam C	Spring 05	143	.094	.986	-.006	1.063
	Fall 04	152	-.089	1.009	.005	.939
Exam D	Spring 05	143	.157*	1.005	-.411	.925
	Fall 04	148	-.152	.080	.395*	.908
Exam E	Spring 05	139	.072	.969	-.003	.968
	Fall 04	145	-.069	1.027	.002	1.033
Exam Total	Spring 05	132	.144*	.983	-.210	.972
	Fall 04	139	-.138	1.000	.200*	.988

Note. *M* = Mean; *SD* = Standard Deviation. *M*s and *SD*s are based on normalized *z*-scores.

*p < .05 positioned by the higher of the two comparison means.

between variable to investigate the potential interaction of semester and exam-item type (notes and readings). A significant interaction was found between the two independent variables ($F = 76.062$, $p < .000$). Specifically, adjusted Sidak pairwise comparisons showed a significant difference between performance on the notes exam items for the spring 05 and fall 04 semesters. That is, students in spring 05 performed significantly better on the notes exam items than students in fall 04 ($p < .05$). The reverse pattern was seen for the non-targeted readings exam items. On these items, students in the fall 04 section significantly outperformed students in spring 05 ($p < .05$). Figure 1 visually illustrates these patterns.

Relationship of Student Performance on Daily-Writing Activities to Exam Performance

Table 3 summarizes the relationships between frequency and quality of student performance on the writing activities and performance on exam items (targeted and non-targeted). Student performance on the targeted multiple-choice exam items across units was correlated with several writing dimensions: total number of writing activities each student engaged in across units out of a possible 25, total-writing score earned on writing activities across units out of 125, and total of students' average-writing scores across units out of a possible 25. To get these composite scores, analogous scores first were computed for each unit and then summed across units.

Overall, composite writing averages correlated most strongly with performance on the targeted notes multiple-choice exam items ($r = .624$, $p < .01$). Although total score on all daily-writing activities also significantly correlated with performance on the

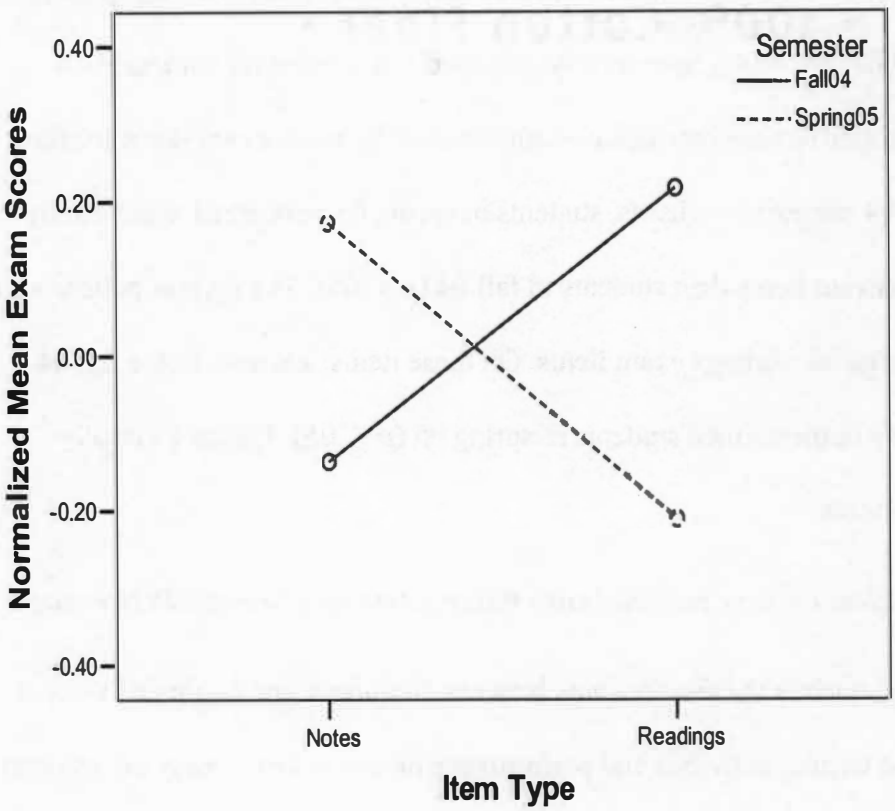

Figure 1. Exam performance on notes and readings items across fall 04 and spring 05 semesters.

Table 3

Correlations between Writing Dimensions and Exam Performance

		Types of exam items		
Writing dimension	*n*	Exam notes	*n*	Exam readings
Writing number	132	.087	133	.136
Writing total	132	.506*	133	.484*
Writing average	131	.624*	132	.573*

Note. .624 is significantly greater than .506 ($t = 3.12$, $p < .001$)

* $p < .01$.

targeted exam items (r = .506, p < .01), the writing-averages variable was a stronger correlate, t (131) = 3.12, p < .001. Student performance on the targeted exam items was not significantly related to the number of writing activities students engaged in (frequency). Composite writing-average scores and total score for all writing activities also were significantly related to performance on the non-targeted exam items (r = .573, p < .01, r = .484, p < .01, respectively), though the magnitude for these correlations was somewhat lower than the correlations with the targeted items.

A multiple-regression analysis (using the three writing-activity dimensions as potential predictors of exam performance) further showed that writing-average was the only writing-activity dimension that significantly predicted performance on the targeted notes exam items (r^2 = .385). That is, student's total of all unit averages on the writing activities accounted for 39% of the variance in exam performance on the targeted items. Similarly, writing average was the only writing dimension that significantly predicted performance on the non-targeted exam items (r^2 = .323). Hence, total writing-average was subsequently compared with other known predictors of exam performance in the course.

Relationship of Writing Averages, Critical Thinking, Vocabulary, and Attendance to Exam Performance

A correlation matrix was computed to determine how well the writing averages, compared to critical thinking, pre-vocabulary, and attendance correlated with exam performance on exam items. Table 4 shows that writing average was more strongly correlated with exam performance on the targeted notes items (r = .624, p < .01) than was

Table 4

Correlations between Course Variables and Exam Performance

Course variable	*m* (sd)	*n*	Exam notes	*n*	Exam readings
WGCTA	26.7 (5.6)	128	.469*	128	.376*
Pre-vocabulary	27.9 (8.2)	130	.505*	130	.397*
Attendance	21.7 (3.7)	132	.082	133	.142
Writing average	18.0 (3.5)	131	.624*	132	.573*

Note. For exam notes, .624 is significantly greater than .469 ($t = -2.03$, $p = .02$), .505 ($t = -1.61$, $p = .055$), and .082 ($t = -6.36$, $p < .001$). Likewise, for exam readings, .573 is significantly greater than .376 ($t = -2.33$, $p = .01$, .397 ($t = -2.09$, $p = .02$), and .142 ($t = -5.01$, $p < .001$).

*$p < .01$.

critical thinking, pre-vocabulary, or attendance. Statistical comparisons of these correlations showed that the correlation for writing-average was significantly greater than for the correlations involving the WGCTA, pre-vocabulary, and attendance. Although the correlations were lower, a similar pattern of correlations held for the exam-readings items.

To further evaluate the predictive linkage of writing average to targeted exam performance, stepwise multiple-regression analyses compared the predictive potential of writing average with pre-course critical thinking, pre-course vocabulary, and attendance. Table 5 shows that average performance on the writing activities was a better predictor of performance on both exam-notes and exam-readings scores than was vocabulary, critical thinking, or attendance. Although vocabulary and critical thinking added to the adjusted R Square, writing-average alone accounted for 37% of the variance in targeted exam items. Writing average also was the strongest predictor of performance on the readings-exam scores, accounting for 32% of the variance. Vocabulary added to the adjusted R Square for the non-target items but critical thinking did not. Total attendance did not enter the prediction equation for either exam-notes or exam-readings scores.

Unit Writing-Average Scores and Performance on Exam Items

Table 6 shows that average-writing scores in each unit significantly correlated with performance on both the exam-notes and exam-readings scores. Correlations with the exam-notes scores ranged from .378 for Unit D writing-average scores to .604 for Unit E average scores, whereas correlations with exam-readings scores ranged from .386 (Unit D) to .536 (Unit C). Stepwise multiple-regression analyses were conducted to

Table 5

Stepwise Multiple Regression for Potential Predictors of Exam-Notes and Exam-Readings Scores

	Model Summary for R Square Adjusted	
Course variable(s)	Notes	Readings
Writing average	.374	.315
Writing average and vocabulary	.529	.407
Writing average, vocabulary, and critical thinking	.546	

Note. All correlations are significant ($p < .01$). Writing average represents the total of all unit writing averages. Attendance total did not enter the regression equation for either exam-notes or exam-readings scores.

Table 6

Correlations between Unit Writing-Average Scores and Exam Performance

Unit writing average	*n*	Exam notes	*n*	Exam readings
Unit A writing average	132	.470	133	.408
Unit B writing average	132	.423	133	.394
Unit C writing average	132	.568	133	.536
Unit D writing average	132	.378	133	.386
Unit E writing average	131	.604	132	.499

Note. All correlations are significant ($p < .01$).

determine specifically which unit(s) best predicted exam performance on the notes items and then on the readings items. Results showed that students' average scores on Unit E writing activities best predicted performance on the composite targeted exam items (Adjusted r^2 = .360). Average scores on Unit C writing activities added a significant amount of predictability for the targeted exam items (Adjusted r^2 = .446). Unit C average scores best predicted performance on the non-targeted reading items (Adjusted r^2 = .286). Likewise, the addition of Unit E writing activity averages increased the predictability to 35%.

Relationship of Each Unit's Writing-Average Scores to Composite Writing-Average Scores

Each unit's writing-average scores correlated significantly ($p < .01$) with the composite writing-average scores: Unit A = .757, Unit B = .811, Unit C = .812, Unit D = .708, and Unit E = .705. To determine how well each unit's writing-average scores predicted the composite writing-average scores, stepwise regression analyses were conducted using the unit averages as predictors and the total averages as the criterion variable. Table 7 shows Unit C writing-average scores to be the strongest predictor of the composite writing-average scores (Adjusted r^2 = .656). However, adding Unit B average scores to Unit C average scores increased the predictability to 84%. The writing-average scores for the remaining units also contributed significantly to the composite writing-average scores, but to a lesser degree than the scores for Units C and B.

Table 7

Summary of Stepwise Regression Analysis for Prediction of Composite Writing Averages from Unit Writing Averages

Model Summary	
Units	R Square Adjusted
Unit C	.656
Unit C and Unit B	.840
Unit C, Unit B, and Unit D	.916
Unit C, Unit B, Unit D, and Unit A	.959
Unit C, Unit B, Unit D, Unit A, and Unit E	1.000

Note. All correlations are significant ($p < .01$).

Relationship of Each Day's Writing Average Score to Composite Day Writing Average Score

To compute writing averages for each day within units, students' actual scores (1-5) on each day's writing activities were summed across units. This sum was divided by the number of writing activities the student attempted for each day across units. For example, if a student earned a 2 on *Day 1* of Unit A, a 3 on *Day 1* of Unit B, did not complete *Day 1* in Units C or D, and earned a 4 on *Day 1* of Unit E, the average *Day 1* score would be 2 plus 3 plus 4, divided by 3 = 3. Each student's composite writing-average score across days was then computed by adding average scores across days, each of which had been computed across units. For example, a student who missed class for all of Unit C would not have a writing-average score for unit C or a composite writing-average score across units (which required a writing-average score for each unit). Nevertheless, this student could still have a writing average for each day across units by completing at least one writing activity for each of the five days across units. Thus, composite writing averages based on days could be slightly different from composite writing averages based on units.

Pearson correlations were computed to determine how performance on each day's writing-averages across all units related to composite writing averages for all days across units. All of these correlations were significant ($p < .01$), ranging from .725 on Day 1 to .867 on Day 2; correlations for the remaining days were .825 for Day 3, .816 for Day 4, and .812 for Day 5. Multiple-regression analyses were conducted to examine how each day's writing averages predicted the composite writing averages across days.

Table 8 indicates that Day 2 averages proved to be the strongest predictor of the composite writing averages (Adjusted $r^2 = .727$), accounting for 73% of the variance in the composite scores. However, scores on all days added to the prediction of the composite average scores across all days.

Relationship of Daily Writing Average Scores and Composite Day Writing Averages to Exam Performance

Days 1 through 5 writing-average scores and the composite day writing averages were correlated with exam performance (see Table 9). Findings showed all of these correlations to be significant at the .01 level, with the exam-notes correlations across days ranging from .448 for Day 1 to .562 for Day 5 and exam-readings correlations across days ranging from .406 for Day 1 to .516 for Day 5. As expected, the composite day scores correlated more highly with both exam measures (.625 for notes items and .575 for readings items) than did any of the individual day correlations. Although somewhat lower than the correlations with the exam-notes scores, the correlations with exam-readings scores yielded a pattern of relationships comparable to that for the exam-notes scores. Next, multiple-regression analyses were conducted to determine how well each day's writing activity average score would predict exam performance on the notes and readings exam items. Day 5 writing-average scores were found to be the best single day predictor of performance on the targeted notes exam items (Adjusted $r^2 = .31$). Adding Day 1 and Day 4 averages increased the predictability to 38% on notes exam items. For non-targeted readings items, averages on Day 5 plus Days 2 and 1 provided the best predictability, accounting for 31% of the variance in exam performance on the reading items. Finally,

Table 8

Summary of Stepwise Regression Analysis for Prediction of Composite Daily-Writing Averages from Daily-Writing Averages

Model Summary	
Day averages	R Square Adjusted
Day 2	.727
Day 2 and Day 5	.870
Day 2, Day 5, and Day 1	.930
Day 2, Day 5, Day 1, and Day 3	.974
Day 2, Day 5, Day 1, Day 3, and Day 4	1.000

Note. All correlations are significant ($p < .01$).

Table 9

Correlations between Daily and Composite-Day Writing Averages and Exam Totals

	Correlations			
Daily and composite	*n*	Exam-notes total	*n*	Exam-readings total
Day 1	132	.448	133	.406
Day 2	132	.511	133	.476
Day 3	132	.487	133	.464
Day 4	132	.535	133	.458
Day 5	132	.562	133	.516
Composite days	132	.625	133	.575

Note. All correlations are significant ($p < .01$).

multiple-regression analyses were conducted to determine how well each day's writing average scores versus composite writing-average scores across days predicted performance on the targeted notes and non-targeted readings exam items. Results show that composite day writing averages was the only significant predictors of exam performance on targeted items (Adjusted $r^2 = .387$), accounting for approximately 39% of the variance in these exam scores. Similarly, composite day writing averages was the strongest predictor on the non-targeted items, accounting for 33% of the variation in performance.

Differences in Writing Averages across Days and Across Units

Tests of within-subjects effects and pairwise comparisons were used to assess the variations in student performance on the writing activities across days for the treatment group (Spring 05). Repeated-measures analysis showed a significant within-subjects effect across days, $F(3.341) = 21.312, p < .001$. Pairwise comparisons using Sidak adjustment revealed that student averages for Day 1 writing activities ($M = 3.21$, $SD = 1.09$) differed significantly ($p < .05$) from Day 3 ($M = 3.66$, $SD = .896$), Day 4 ($M = 3.69$, $SD = .927$), and Day 5 averages ($M = 3.74$, $SD = .787$). Also, Day 2 averages ($M = 3.34$, $SD = .992$) differed significantly from Day 3, 4, and 5 averages. In general, students' average writing scores increased from Day 1 to Day 5 (see Figure 2).

Findings also indicated a significant within-subjects difference across units, $F(3.711) = 5.220, p < .001$), for writing-average scores in the treatment group (Spring 05). Overall, student averages were lowest on the Unit B writing activities ($M = 3.45$, $SD = 1.00$) and highest on the Unit A writing activities ($M = 3.78$, $SD = .957$). Mean writing-

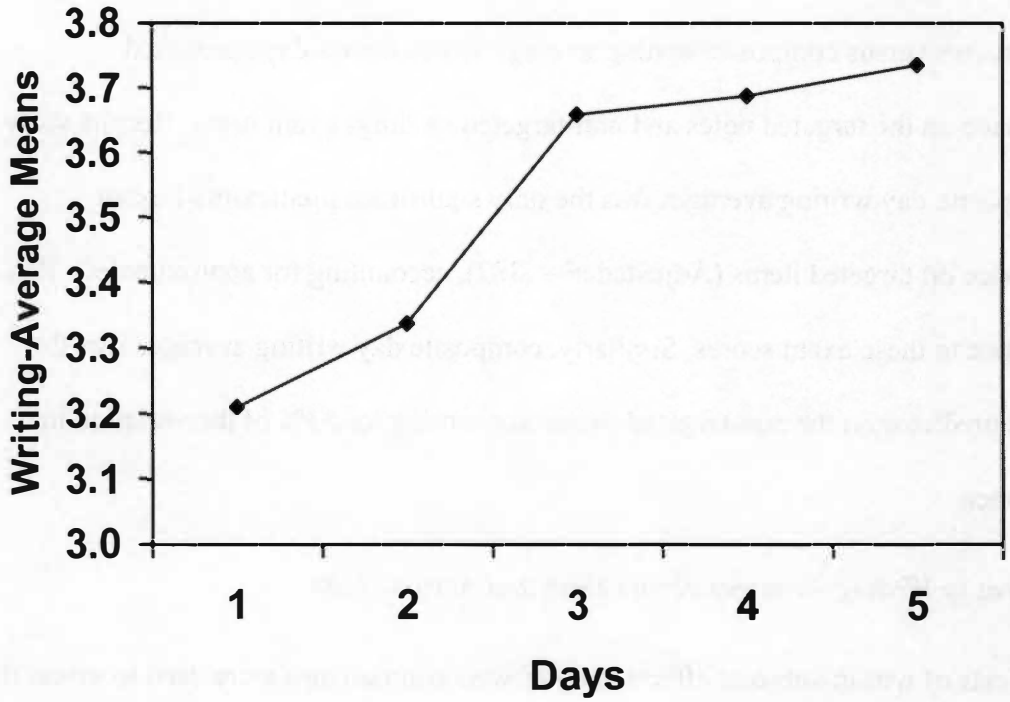

Figure 2. Writing-average means across days.

average scores for all units were 3.78 for Unit A, 3.45 for Unit B, 3.66 for Unit C, 3.53 for Unit D, and 3.54 for Unit E. Standard deviations ranged from .82 for Unit C to 1.00 for Unit B. Pairwise comparisons revealed significant differences ($p < .05$) between writing-average means for Units A and B, and Units B and C (see Figure 3).

Differences in Exam Performance for High and Low Writing-Average Performers

To further determine the effect of choosing to engage in the daily writing activities, two performance groups were formed based on students' writing averages. Students whose writing averages were in the bottom quartile were classified as low performers and students whose writing averages fell in the top quartile were classified as high performers. Table 10 provides a summary of normalized targeted exam scores for high and low writing-average performers. The low writing-average performers mean score on targeted exam notes items was -.701 ($SD = .974$), compared to .816 ($SD = .694$) for high writing-average performers. An independent samples t-test revealed a significant between-groups difference in the notes-items on the exams, $t(61) = 7.174$, $p < .000$. That is, students in the high writing-average group performed significantly better on the targeted notes exam items than students in the low writing-average group. Similar results were found for the non-targeted reading exam items. That is, students in the low writing-average group performed significantly lower on the non-targeted reading exam items, $t(61) = 5.986$, $p < .000$, than students in the high writing-average group.

The effect size for the difference between mean scores on the notes versus readings items was greater for the high writing-average group than for the low group.

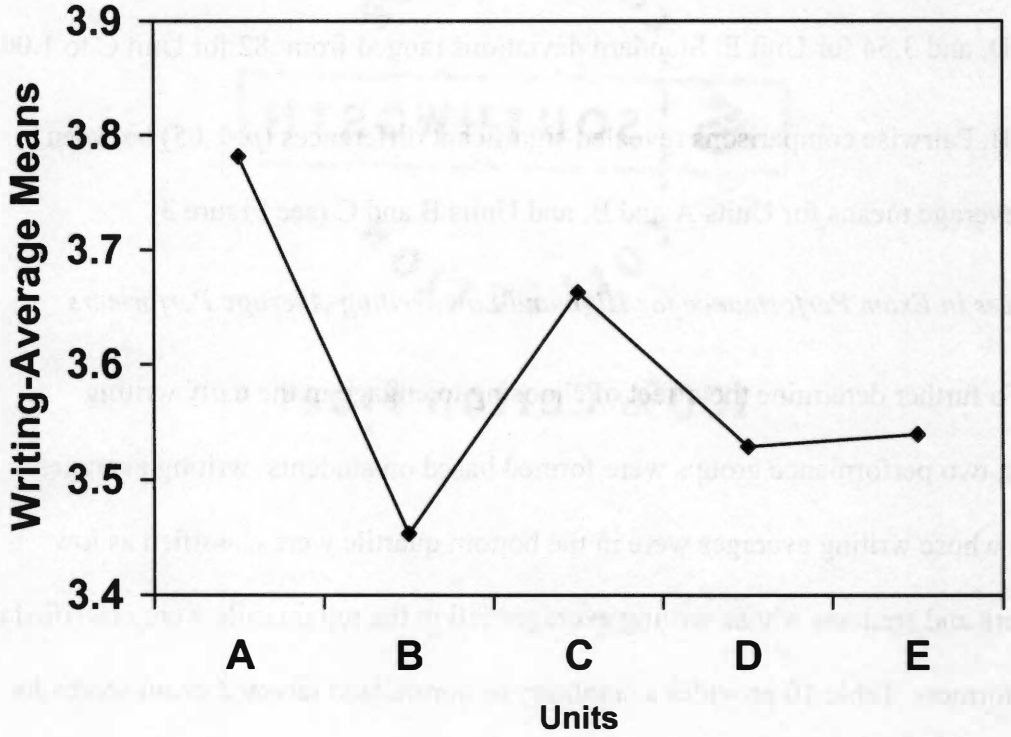

Figure 3. Writing-average means across units.

Table 10

Normalized Exam Scores for High and Low Writing Performers

	Group	*n*	*M*	*SD*
Exam-notes total (*z*-score)	Low	30	-.701	.974
	High	33	.816	.694
Exam-readings total (*z*-score)	Low	30	-.953	.992
	High	33	.390	.785

Note. High and low writing performers differed significantly ($p < .01$) for both exam-notes items and exam-readings items.

That is, the effect size for target versus non-target exam scores for the low writing-average group was -.254, compared to .543 for the high writing-average group. Consequently, students who performed higher on the writing activities not only performed better on the targeted exam items, they also showed a greater difference in their performance on the two types of exam items (targeted versus non-targeted).

Chapter IV

Discussion

This study was based on the premise that by implementing a technique designed to increase pre-class preparation, student subsequent exam performance would consequently increase. Researchers have documented the discrepancy between the amount of time instructors recommend students spend preparing for class and the actual time students spend (www.iub.edu/~nsse). Accordingly, any technique that encourages student preparation would appear advantageous. Likewise, an instructional technique that encourages deep learning of the course content should facilitate the use of deep learning strategies in preparing for major exams.

The findings of this study point to the following conclusions: (a) providing daily writing activities that involve comparison of major concepts in instructor notes appears to improve performance on multiple-choice exam items related to those notes but not on exam items unrelated to those notes (items based on course readings); (b) the average quality of writing performance is a better predictor of exam performance than is the frequency of writing experiences or total credit on the writing experiences irrespective of their frequency; (c) average-writing performance on instructor-notes concepts better predicts performance on multiple-choice exams than does pre-course critical thinking, pre-course vocabulary, and attendance; (d) writing averages scores by unit significantly predict exam performance, but not as well as composite writing performance across all units; (e) performance on each day's writing activity within units significantly predicts exam performance, but not as well as composite writing performance across days. (f)

performance on each day's writing activity tends to improve across days within course units; and (g) high writing-activity performers significantly outperformed low performers on unit exams.

A major finding from the current study was the differential performance of students during the fall 04 (no treatment) and spring 05 (treatment) semesters. As hypothesized, students did significantly better on the notes exam items during the treatment semester. Although there is no empirical validation of student's out-of-class preparation, a reasonable assumption is that the increased exam performance on the notes items resulted from the out of class preparation for the writing activities. Additionally, students could have procrastinated less during the treatment semester. Because students were engaging in daily credit-producing writing activities, they may have been more inclined to prepare for the posted conceptual pairs each day. Alternately, students in the non-treatment semester may have been disinclined to read their instructor notes before coming to class. One possibility is that although students in the non-treatment semester may have looked over the instructor notes, their pre-class preparation may have been minimal compared to students in the spring 05 semester who were preparing for the writing activities. Students in the non-treatment semester were probably more inclined to study their instructor notes right before the exams. Alternately, students in the treatment semester were more likely to study the notes on a regular basis as they prepared for the daily writing activities. According to research on the spacing effect (Grote, 1995; Seabrook, Brown, & Solity, 2005), spaced repeated practice leads to better performance than massed practice. Perhaps students' study patterns for exams differed across the two

semesters. That is, students in the treatment semester may have adopted more of a distributed practice approach than those in the non-treatment semester.

The nature of the writing activities could have led to differential performance on the exam-notes items across semesters. Specifically, for success on the writing activities, students were required to synthesize and integrate information from multiple sources. Students were required to infer relationships between important course concepts. Typically, the link or similarity between posted concepts was not explicitly indicated in the notes or power point slides; instead, students were to infer a commonality based on the given information. In this way, the writing activities facilitated student's use of complex learning strategies. Potentially, this daily practice with complex learning strategies may have increased the use of such strategies on the unit exams.

The pattern of exam performance was opposite for the non-targeted readings exam items. Specifically, students in the non-treatment semester did significantly better on the readings items than students in the treatment semester. One probable explanation for this finding is the differential amount of class time devoted to discussing the out of class reading material. That is, in fall 04, a short in-class review was devoted to the discussion of the reading materials. Alternately, in the spring 05 semester, no class time was devoted to discussion of the out of class readings. Although students were told to direct their readings questions to instructors via phone or email, anecdotal evidence suggested students rarely took advantage of this opportunity. Typically, there was no instructor-student discussion of the readings. Results support the hypothesis that devoting some class time to the discussion of out of class reading will improve subsequent exam

performance on items pertaining to the readings. Another possible explanation for the differential finding is that during the non-treatment semester, students devoted more time to the out of class readings and less time reviewing the instructor notes. Conversely, during the spring 05 semester, students may have devoted less time to the out of class readings because they were preparing for the daily writing activities. Inasmuch as the study involved the simultaneous addition of the daily writing activities and the deletion of the reading review during the treatment semester, future research should address the potential impact of each procedure separately.

The superiority of writing average scores in predicting exam performance is another major finding from the study. Specifically, compared to writing number and writing total, results showed that students' writing average was the only writing-activity dimension to significantly predict performance on the exams. Although writing total was strongly correlated with exam performance on notes and readings items, the relationship was stronger between exam performance and writing average. Interestingly, student performance on exams was not related to the frequency of engagement in the writing activities. That is, the total number of writing activities students engaged in was unrelated to performance on either notes or readings exam items. Thus, it was not the students' engagement in the writing activities (frequency) per se that predicted exam performance; rather, students' quality of engagement in the writing activities significantly predicted success on the exams.

Another interesting twist was that writing average was almost as strongly related to readings exam items as it was to notes items. Given the direct linkage between writing

activities and notes exam items, it was expected that the writing activities would be a stronger predictor of performance on notes rather than readings items. However, findings show that writing averages significantly predicted performance for both exam item types (notes and readings). Several explanatory factors could account for this finding. First, writing averages may have been related to a general achievement pattern. That is, preparation for the daily writing activities may have been accompanied by careful attention to the out-of-class readings.

Another possibility is that the writing activities equipped students with various strategies for answering difficult multiple choice exam items, regardless of item type (i.e., notes and readings). As such, strategies learned for use on the daily writing activities could have facilitated the use of such strategies on all exam items, not just those items pertaining to information in the instructor notes. For example, students could have learned to infer relationships between concepts by pulling information from multiple sources. This new strategy could, in turn, enhance student's ability to make inferences on the more difficult exam items, both notes and readings.

Another critical finding from the study was the differential relationships between writing average, critical thinking, vocabulary, and attendance with exam performance. Findings clearly show that writing average was more strongly related to notes exam performance than was critical thinking, vocabulary, or attendance. Similarly, writing average was the strongest correlate with readings exam items. Inasmuch as both the writing activities and exams are tied specifically to the course content, it logically follows that performance on the writing activities would be more strongly related with exam

performance than other, less directly related variables. Interestingly, in evaluating the regression analyses, total attendance did not enter the prediction equation for either exam item type. This result resembles the aforementioned finding that writing frequency did not predict performance on exams. In both cases, the findings suggest that simply attending class and engaging in the writing activities is ineffectual in improving exam performance. Instead, the student's high-level preparation and engagement in the writing activities facilitates higher performance on exams. Because student's average on the writing activities was a significant predictor of exam performance, poor performance on the writing activities could serve as a "red flag" for instructors. That is, instructors could offer instructional support to students whose performance on the first few writing activities is significantly below that of other students.

Another research question was whether scoring only some of the writing activities would decrease the instructor's ability to predict exam performance. In the main, while the composite scores proved to be the best overall predictor of exam performance, unit and daily averages were also found to be significant predictors of exam performance. Although it is unclear what the effect would have been on student writing and exam performance if students had earned credited on each day's writing activity versus randomly selecting only one day to credit, it is possible that the students would have prepared better each day by knowing that every day's writing score would be counted toward their grade. However, knowing that one day's writing score was only one of five writing scores to be credited for a unit might have lessened student effort on any particular day. In contrast, in the current study, each day's writing score potentially could

be a student's writing score for that unit. Thus, doing one's best each day would seem advisable from the student's perspective.

From a strictly statistical perspective, counting only one day's writing performance did not predict exam performance as well as using the composite ratings across all days in units as the predictor. Yet the correlations between each day's writing scores (across units) and exam performance were in the medium to large range, with all correlations highly significant. Although the differential effect of counting every day's writing performance versus one randomly selected day's performance remains a research issue, we would predict that counting each day's writing performance would have much the same impact on student writing and exam performance as counting writing performance on one randomly selected day.

Another interesting caveat in the findings is the variability in student performance on the writing activities across units and days. Basically, student performance on the writing activities increased from Day 1 to Day 5. The most probable explanation for this finding is that writing activity performance was affected by temporal proximity to a unit exam. That is, Day 1 occurs on the class period following an exam and students may have been disinclined to prepare after just studying for the previous exam. Likewise, students were probably more likely to prepare for the Day 5 writing activities because the exam was forthcoming. This pattern of improved performance across days within units appears consistent with the notion of post-reinforcement pause. That is, because unit exam scores provided a major source of reinforcement for students, they may have had a tendency to prepare less for the writing activity on the class sessions immediately following unit

exams. In the cross-unit analysis, student's performed best on the Unit A activities and worst on the Unit B activities. Students may have been more eager to prepare for the writing activities during Unit A, given that it was the first unit in the course. Likewise, students may have been disinclined to prepare for Unit B writing activities because: (a) they realized the writing activities only accounted for a minimal amount of overall course credit or (b) they became busy in other courses and therefore had less time to devote to 210.

To determine the differential effects of high versus low quality performance on the writing activities, students who scored in the bottom quartile on the writing activities were compared with those who scored in the top quartile. In the main, high writing-average performers did significantly better than low performers on unit exams. By computing effect sizes for the difference in exam item type for high and low performers, performance differences in the two item types could be compared. Results indicate a greater effect size for the high performers. That is, the difference in their performance on the target and non-target items was substantially greater than students for in the low performing group. Although they outperformed low performers on both item types (notes and readings), high-performers showed greater difference in their performance on the two types of exam items. Thus, while high-quality performance on the writing activities led to increased overall performance on the unit exams, this elevation in scores was especially pronounced for the targeted notes items.

Although the empirical findings of the study suggest that daily writing activities related to major course concepts can improve multiple-choice exam performance, the

linkage between these writing activities and exam performance may not constitute a direct cause-effect relationship. One finding that points to the possibility of a more indirect explanation is the relationship between scores on the writing activities and performance on non-targeted exam items. Students did not do better (in fact, they did worse) on the latter items in the treatment semester than the non-treatment semester, and scores on the writing activities were less predictive of performance on the non-targeted than the targeted items. Nonetheless, writing scores were significantly and strongly related to performance on the non-targeted items. This finding may suggest that level of preparation for the daily writing activities is part of a larger work-habits scheme that affects performance on all aspects of the course. For example, students who prepared well for the writing activities also may have been more diligent in doing their readings (sources of non-targeted exam items) and answering study questions over the readings than those who were negligent in preparing for the writing activities.

Another way that the daily writing activities could have indirectly affected performance on the exams was through the quality of class discussion. Although not confirmed through systematic observation, the quality of class discussion appears to have been much improved by the writing activities. All instructors in the course during the treatment semester also had worked with the course the non-treatment semester. They were asked to informally identify changes in student behavior during the treatment semester compared to the non-treatment semester. What stood out in their anecdotal feedback was the perspective that class discussion was much more informed during the treatment than the non-treatment semester. All instructors reported that students were

much more prepared to discuss major concepts in the instructor notes and were more accurate in their comments during the treatment semester.

The improved quality of class discussion likely extended the benefits of pre-class preparation for the writing activities. It should be noted that in-class discussion of any particular pair of concepts (e.g., humanism versus behaviorism, intrinsic versus extrinsic reinforcement) always occurred after students had been asked to write on a selected conceptual pair (i.e., discussion of a conceptual pair never preceded writing on that conceptual pair). Additionally, instructors reported a sharp reduction in tardiness during the treatment semester. Because the writing activities occurred during the first five minutes of class, students were probably inclined to arrive on time so that they would not miss the activity.

One might assume that because writing activities occurred on a high percentage of class sessions, with each class session offering the potential for students to earn writing credit, student performance on the exams might have been related to attendance patterns. Yet, of all the potential predictors of exam performance (pre-course critical thinking, pre-course vocabulary, attendance, and average daily writing), only attendance proved to have no predictive potential and average daily writing had the greatest predictive potential. Admittedly, attending class was essential for participating in the daily writing activities, but simply being in class made little contribution to performance on the exams in the course. Nonetheless, the prospect of a daily writing activity that could result in course credit may have motivated some students to attend class when they would otherwise have stayed away.

Although performance on the daily writing activities had stronger predictive potential than did previously established contributors to exam performance (Williams & Eggert, 2002; Williams & Worth, 2002), instructor evaluation of students' daily writing products could be a very labor-intensive process. All writing products were evaluated for research purposes in the current study, but the system employed for official credit required the evaluation of only one writing activity per student per unit. Neither students nor the instructor knew which writing activity would be evaluated for credit until it was randomly selected by a student at the completion of the last writing activity in the unit. Because not all students had attended the class on the selected day, the number of writing activities to evaluate in a particular course unit was usually less than the number of students in the class.

The necessary time to evaluate the writing products was likely affected by the clarity of the rating procedures. Given that the writing activities were rated by a number of different graduate teaching assistants across sections of the course, it was important to have a uniform rating method for all writing products. To facilitate this process, the supervising instructor for all sections developed a prototypic answer for all conceptual pairs, which raters used in judging the answers of students. Plus, both raters and students were told that answers must include at least one commonality and/or linkage between members of the conceptual pair selected for writing each day and at least four differences between the concepts. Incorrect statements of similarities or differences detracted from a student's net credit on a writing activity. The manageability of the rating procedures used

in this study is attested by the consistently high inter-rater reliability for the writing products.

Certainly, both class time and rating time outside of class are important practical issues in appraising the transportability of the treatment package used in this study. The writing activities were done the first 5 minutes of 50-minute class periods for 5 class sessions in each unit, constituting 25 class sessions across all units. Thus, each treatment day, one-tenth of the class time was devoted to the writing activity. For each unit, 25 minutes out of 350 minutes (7 class sessions per unit) were devoted to the writing activities, and for the total course 125 minutes out of 1750 minutes were devoted to the writing activities. The defensibility of using this much class time for any instructional/assessment procedure must be judged by its net effect on important outcome variables in a course (in this case course exams). Certainly, the findings of the current study suggest that time devoted to the writing activities was well spent.

Student responses in the five-minute segments at the beginning of class were seldom more than one page. Plus, raters knew exactly what to look for in student responses. With this arrangement, each student paper could be evaluated in a minute or less. For our large classes, instructors took no more than an hour to evaluate the writing products for a particular day. When one considers that rating the writing products took no more than one hour per seven days of class in a course unit (five writing days and two days devoted to essay quizzes and multiple-choice exams), instructor time-investment in the writing activities appears manageable. Additionally, by improving exam performance, instructors could save time in dealing with the fallout of poor exam performance (e.g.,

responding to student complaints about the exams and meeting with students on an individual basis to go over their exams).

Certainly, the current study should be viewed as an initial investigation in a research agenda devoted to pinpointing what factors contribute to the efficacy of the daily-writing arrangement and finding ways to make the procedure more efficient and effective. I strongly believe that having students actually write on course concepts most every day, with the prospect of receiving credit for their written products, was seminal to the efficacy of the daily writing activities. However, additional research is needed to confirm that the writing per se versus simply studying the concepts or discussing targeted concepts at the beginning of class is most basic to the effectiveness of the writing arrangement. For example, students in all sections of the course could have the same conceptual pairs posted before each class session and be told to study the concepts as a way of integrating specific information in the notes. Beyond the general instructions, sections of the course would receive different orientations as to how they would deal with the conceptual pairs: some sections would be told to study the conceptual pairs and be prepared to discuss any one of them at the beginning of the subsequent class period; other sections would be told to study the conceptual pairs and be prepared to write on any one of them at the beginning of the next class period with no credit contingency; and still other sections would receive the same treatment employed in the current study (i.e., studying the conceptual pairs and being prepared to write on any one of them for possible credit).

If follow-up research should indicate that writing for potential credit proves to be the superior arrangement for promoting performance on exam items, then various credit contingencies could be examined. Although other research (Oliver & Williams, 2005) suggests that rewarding quality of performance on an activity is more potent than rewarding completion of the activity, this differential effectiveness should be confirmed for the writing activities (with some sections given credit for submission of the writing products and others for the quality of their writing products). Assuming that the quality contingency proves more powerful than the completion contingency (as I would expect), researchers should compare several variations of the quality contingency: giving credit for writing products each day they are scheduled, giving credit for the writing products on a randomly selected day in each unit (as was done in the current study), and giving credit to each student based on the writing product of a randomly selected student on a randomly selected day in each unit. Should the latter arrangement prove as effective as the other contingencies, it would represent the most efficient arrangement from the instructor's perspective (requiring the evaluation of only one student's writing product each unit). However, because this contingency has the potential to under-credit students who do well on the writing activities, it should probably be used with bonus credit rather than regular credit.

In addition to exploring different types of credit contingencies, the effects of differential amounts of potential credit should also be examined. In the current study, students could earn five points per unit, amounting to a ratio of about one point's potential credit for each day's writing activity (i.e., students had to write for five days to

maximize their chances for the five credit points). As previously noted, the total potential credit for the writing activities in the course amounted to no more that 4% of the total credit. Our research team has found in research designed to improve students' vocabulary development in the course that doubling the amount of credit offered for vocabulary mastery significantly enhanced their level of vocabulary development (Turner & Williams, 2004). Similarly, offering different amounts of credit for the quality of writing products would be necessary to determine the optimal amount of credit to enhance student mastery of major course constructs.

To maximize proficiency in comparing the exam performance of classes receiving different instructions and credit contingencies relative to the writing activities, two other arrangements need to be added to the research agenda. First, a pre-course exam related to all units in the course should be administered to all sections. Consequently, if sections obtain significantly different scores on the pre-test, those scores could be used as a covariate in examining differences in exam performance across sections. Secondly, examining the quality of class discussion is pivotal to fully evaluating the effects of the different instructions and contingencies linked to the writing activities. As previously noted, one way writing activities could improve exam performance is by improving the quality of class discussion. However, that presumed linkage in the current study was supported only by anecdotal feedback from course instructors.

Evaluating the quality of class discussion would be far more difficult than the assessment of the variables targeted in the current study. Such an evaluation would likely require videotaping class discussion to verify that instructor input was similar across

sections and to appraise the accuracy of student comments in class. Only raters who were intimately acquainted with the content of the course could accurately make those judgments. The time investment in making ratable tapes and then rating instructor and student comments in class would be considerable. Though difficult, such an analysis would be invaluable to the line of research begun in the current study.

The goal of continuing research in this area is to develop highly efficient and effective ways to improve student preparation for class each day, their discussion in class, and their performance on major course assessments (such as principal exams in a course). The current study is a significant step in that direction but certainly not the final word.

LIST OF REFERENCES

LIST OF REFERENCES

Center for Postsecondary Research. (2004). *National survey of task engagement.* Annual Report. Retrieved April 4, 2005 from www.iub.edu~nsse.

Connor-Greene, P. A. (2000). Assessing and promoting student learning: Blurring the line between teaching and testing. *Teaching of Psychology, 27*(2), 84-88.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58,* 438-481.

DeRoma, V. M., Young, A., Mabrouk, S. T., Brannan, K. P., Hilleke, R. O., & Johnson, K. Y. (2003). Procrastination and student performance on immediate and delayed quizzes. *Education, 124*, 40-48.

Dickie, L. O. (2003). Approach to learning, the cognitive demands of assessment, and achievement in physics. *The Canadian Journal of Higher Education, 33*, 87-111.

Fulkerson, F. E., & Martin, G. (1981). Effects of exam frequency on student performance, evaluations of instructor, and text anxiety. *Teaching of Psychology, 8,* 90-93.

Grote, M. G. (1995). Distributed versus massed practice in high school physics. *School Science and Mathematics, 95*, 0036-6803.

Higher Education Research Institute. (2001). *The American freshman: National norms for fall 2001.* Retrieved on April 5, 2005 from http://www.gseis.ucla.edu/heri/norms_pr_01.html.

Kerkman, D. D., Kellison, K. L., Pinon, M. F., Schmidt, D., & Lewis, S. (1994). The quiz game: Writing and explaining questions improve quiz scores. *Teaching of Psychology, 21,* 104-106.

Martin, F., & Saljo, R. (1976). On qualitative differences in learning: Outcome and process. *British Journal of Educational Psychology, 46,* 115-127.

McDougall, D., & Granby, C. (1996). How expectation of questioning method affects undergraduates' preparation for class. *Journal of Experimental Education, 65,* 43-55.

Seabrook, R., Brown, G. D. A., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19,* 107-122.

Shaw, V. N. (2001). Training in reading skills: An innovative method from college instruction. *Reading Improvement, 38,* 188-192.

Simkins, S., & Maier, M. (2004). Using just-in-time teaching techniques in the principles of economics course. *Social Science Computer Review, 22,* 444-456.

Spies, A. R., & Wilkin, N. E. (2004). Effects of pre-class preparation of legal cases on in-class performance. *American Journal of Pharmaceutical Education, 68*(2), 1-5.

Ruscio, J. (2001). Administering quizzes at random to increase students' reading. *Teaching of Psychology, 28,* 204-206.

Tai-Seale, T., & Thompson, S. B. (2000). Assigned conversations. *College Teaching, 48,* 755-756.

Thome, B. M. (2000). Extra credit exercise: A painless pop quiz. *Teaching of Psychology, 27,* 204-205.

Tuckman, B. W. (1996). The relative effectiveness of incentive motivation and prescribed learning strategy in improving college students' course performance. *Journal of Experimental Education, 64*(3), 197-210.

Tuckman, B. W. (1998). Using tests as an incentive to motivate procrastinators to study. *Journal of Experimental Education, 66*(2), 141-147.

Turner, H. C., & Williams, R. L. (2004). Generic vocabulary development as a predictor and outcome variable in a large human development course. Paper presented at the Mid-South Educational Research Association Meeting, Gatlinburg, TN.

Wallace, M. A., & Williams, R. L. (2002). Multiple-choice exams: Explanations for student choices. *Teaching of Psychology, 30*, 136-138.

Watson, G. B., & Glaser, E. M. (1994). *Watson-Glaser Critical Thinking Appraisal – Form S Manual.* San Antonio: The Psychological Corporation.

Williams, R. L., & Worth, S. L. (2002). Thinking skills and work habits: Contributors to course performance. *The Journal of General Education, 51*, 200-225.

APPENDICES

APPENDIX A

Samples of Conceptual Pairs with Prototypic Answers for Use in Scoring

UNIT A INSTRUCTOR-NOTES CONCEPTUAL GROUPS

Prototypic Answers

Professional journals and publicly supported health/safety agencies—Both sources provide generally credible and recent research information regarding health/safety issues. However, because articles submitted to professional journals are rigorously reviewed by experts in the field, these journals are typically regarded as the most credible and recent source of research information. Reports released by government health and safety agencies (e.g., CDC) usually are conducted by scientists who work for these agencies, yet higher levels of government sometimes exercise editorial prerogatives in removing or revising information considered to be politically sensitive in the reports generated by the agencies.

Correlational and experimental-control group research—Both represent statistical ways of representing relationships between variables. Correlations indicate the quantitative relationship between two variables but without any stipulation of cause and effect. Only the results of experimental-control studies can be used to infer cause-effect relationships between variables. Correlations can range from 0 to + or – 1.00, with the larger values indicating strong relationships and the + or – sign indicating the direction of the relationship. In experimental-control group research, random selection of participants and the use of placebos for control comparisons are important dimensions.

Girls' and boys' drug use—Both males and females appear to use the same types of drugs, but their patterns of use differ by gender. More males than females have used a plurality of tobacco products, alcohol, and marijuana in the last thirty day days than have females. Smokeless tobacco and cigar use is much less common among females than males. However, more females than males now smoke in most developed countries, although gender rates for smoking are about equal among American teens.

Girls' and boys' reported health problems—Adolescent girls report more health problems than do boys. One of the most dramatic gender differences relates to HIV, in which case nearly two-thirds of adolescents infected with HIV are girls. Women become addicted to alcohol more quickly, develop lung cancer at an earlier age, and are more likely to have excessive body fat than do males. Despite the directionality of these gender differences, women tend to live longer than men.

Effects of health education and physical education—Both are likely to improve children's health. Health education reduces onset of smoking, prevalence of obesity, and use of three major drugs in adolescence. Physical education also may contribute to children's intellectual development. Increased gym activity may increase exam scores. Regular physical activity may trigger the growth of new brain cells.

Effects of seatbelts and air bags—Both reduce the likelihood of fatalities in car accidents. Seatbelts reduce the likelihood of death in auto accidents by nearly 50%. Yet

40% of children don't use seatbelts on a regular basis. Air bags provide the greatest protection when seatbelts are buckled. Most individuals killed by air bags were not wearing seatbelts.

Patterns of drug use in late 1970s and 2001—Use of most drugs peaked in mid to late 1970s and then generally declined to the early 1990s, increasing again in the early to mid-90s. However, increase in the 90s did not approximate the peak in the 70s. 2001 data show a moderate decrease in the general use of drugs and most especially for cigarettes.

Comparative use of alcohol and tobacco among teenagers—Alcohol and tobacco represent the most commonly used drugs in adolescence. Alcohol is the drug used by the highest percentage of teenagers (close to 80%) and tobacco is the most widely used drug on a daily basis among teenagers. More than 40% of U.S. high school students currently use (in last 30 days) one or more tobacco products. Alcohol and tobacco can both serve as gateway drugs to the use of illicit drugs, but tobacco is the principal gateway drug leading to the use of other drugs.

Effects of smoking on hearts and lungs—The heart and lungs represent the two body organs most commonly and seriously affected by smoking. Smokers have twice the probability of developing heart disease and three times the probability of dying from heart disease as non-smokers. Smoking raises LDL, lowers HDL, increases risk of blood clot, and pits the lining of the arteries, causing more plaque to stick to the lining of the arteries. Smoking increases the risk of lung cancer by ten-fold. One and one-half times as many smokers die of lung cancer as heart disease.

Parental and peer influences on smoking—Both parents and peers strongly influence whether youth will smoke. Children with parents who smoke are almost 3 times as likely to smoke as children whose parents do not smoke. However, smoking is not as strongly related to family influences as several decades ago. Spending time with peers who smoke is one of the greatest contributors to starting smoking and one of the greatest deterrents to stopping smoking.

Interactive and lecture approaches to drug education—Both of these instructional arrangements are commonly used in drug education. Working together to achieve personally meaningful outcomes may lead to peer bonding counter to the use of drugs as a vehicle of bonding. Lecture approach that attempts to define the reality of drug use for youngsters is less likely to be effective than one in which students can share experiences, ask questions, and question the conclusions of the program leader.

Short-term and long-term physical consequences of smoking—Some short-term consequences, such as respiratory illness, foretell the possibility of long-term life-threatening consequences. Short-term consequences may be more effective deterrents to smoking among young people than the long-term perils of smoking. Short-term consequences include more respiratory illness, yellowing of teeth, greater shortness of breath in exercise, and short periods of energy followed by depression and fatigue. Long-

term, smoking is the number one cause of premature death in the U.S., greatly increasing the probability of terminal respiratory and cardiovascular illness.

Self-directed and cold-turkey approaches to smoking cessation—Both are commonly used methods for trying to stop smoking. Self-directed is a planned sequence of steps designed to reduce smoking cues in one's physical and social environment and increase activities counter to smoking. This approach usually includes setting a target date for quitting, reducing socialization with smokers, developing a non-smoking support group, engaging in regular exercise, and acquiring alternative stress-reduction strategies. The cold-turkey approach involves attempting to give up smoking immediately and completely without making any changes in one's physical and social environment or other habits.

LDL and HDL—Both are forms of cholesterol, with LDL being considered the bad cholesterol and HDL the good cholesterol. LDL adheres to lining of the arteries (atherosclerosis) and HDL tends to cleanse the blood stream of LDL. LDL is more affected by diet than HDL, which is more affected by exercise. Ideally, the ratio between LDL and HDL should be no more than 2 to 1.

Saturated fat, unsaturated fat, and trans-fatty acids—All are fats commonly available in food products. Saturated fats mainly come from animal products and contribute to atherosclerosis, whereas unsaturated fats (polyunsaturated and monounsaturated) may help lower bad cholesterol and overall cholesterol level. Trans-fatty acids result from the partial hydrogenation of vegetable oils and are even more likely to contribute to unhealthy cholesterol levels than is saturated fat.

Young children's and high schoolers' dietary behaviors—Eating habits generally get worse as children get older, with four times as many preschoolers having a good diet than do high schoolers. High school students consume too much salt and red meat, too much saturated fat, and too little whole grain products.

Original food pyramid and redesigned food pyramid—Both emphasize whole grains, vegetables, and fruits. The major differences pertain to plant oils, starches, and red meat. The redesigned food pyramid puts plant oils at the co-base of the pyramid, whereas the original food pyramid placed them at the top of the pyramid. The original food pyramid put starches with whole grains at the base of the pyramid (making no distinction in types of complex carbohydrates), whereas the redesigned food pyramid puts starches at the top of the pyramid. The original food pyramid put red meat with other meat products (e.g., chicken, seafood), whereas the redesigned food pyramid puts red meat at the top of the pyramid.

Yo-yo dieting and healthy dietary intake—Both dietary approaches can affect weight loss and weight gain. Yo-yo dieting involves going on and off of low-calorie diets. When one goes off of a low-calorie diet, the person is likely to regain all the weight lost plus extra weight. The weight regained is likely to have a higher percentage of body fat than

the weight lost. Each time one goes off a low-calorie diet, the person will find it increasingly difficult to lose weight on other low-calorie diets. In contrast, a healthy diet is one that can be maintained on a permanent basis. It has an adequate number of calories coming mainly from whole grains, vegetables, vegetable oils, and fruit. It has lots of fiber and minimal saturated fats and starches.

Body fat and body weight—Both reflect one's fitness level, but body fat is a much better measure of fitness than is body weight. One can technically be overweight but have a low percentage of body fat (weight lifters, for example). Body fat is inversely proportional to amount of lean tissue. Thus, some individuals who get little exercise can be of moderate weight but have minimal lean tissue and excessive body fat.

Endurance, strength-building, and stretching—All three are exercise categories that contribute to fitness, with endurance being the most vital of the three (contributes to cardiovascular health). However, the other two categories contribute to a strong and flexible body, which should minimize injuries in endurance activities (e.g., running and other forms of aerobic activity).

Genital herpes, syphilis, and HIV—All three can be sexually transmitted. HIV is the only one that is both incurable and deadly. Genital herpes is incurable but typically does not present a threat to life. On the other hand, syphilis can be cured if treated in a timely manner, but can be deadly if left untreated.

Abstinence-only and abstinence-plus programs—Both approaches emphasize that abstinence is the only completely effective protection against unwanted pregnancy and STDs. However, abstinence-plus programs also present contraception as an avenue for birth control and protection against STDs. Abstinence-only programs either don't address contraception or portray premarital sex as unacceptable under any circumstances. Abstinence-only programs tend to have a more moralistic emphasis than abstinence-plus programs. Abstinence-only programs appear not to result in a significant delay in the initiation of sexual activity, and abstinence-plus programs have not been shown to increase sexual activity.

APPENDIX B

Conceptual Pairs for Units A, B, C, D, and E

INSTRUCTOR-NOTES CONCEPTUAL PAIRS

For each day except the day of the readings quiz and the day of the unit exam, we will ask you to define and compare the concepts embedded in one of the conceptual groups listed below. Your instructor will select the conceptual group for each day based on the expected scope of discussion that day. You will be given five minutes at the beginning of the class period to write your response to the selected conceptual group. Although you will not be permitted to use your notes in class when you respond to the selected conceptual group, good preparation for this activity would be for you to construct a written response to each group on an out-of-class basis. Your response to each conceptual group should address the following questions: What do the concepts have in common? How are they linked (i.e., are they correlated or does one affect the other in some fashion)? How are they different? This out-of-class preparation will require a close examination of the notes pertaining to the targeted concepts. Put as much detail as you can in your five-minute answer in class. On the last day of the unit that we follow this procedure, we will randomly select one of the days in the unit for which to count your response for credit.

UNIT A INSTRUCTOR-NOTES CONCEPTUAL PAIRS

Professional journals and publicly supported health/safety agencies
Correlational and experimental-control group research
Girls' and boys' drug use
Girls' and boys' reported health problems
Effects of health education and physical education
Effects of seatbelts and air bags
Patterns of drug use from the mid-1970s to 2002-2004
Comparative use of alcohol and tobacco among teenagers
Effects of smoking on hearts and lungs
Parental and peer influences on smoking
Interactive and lecture approaches to drug education
Short-term and long-term physical consequences of smoking
Self-directed and cold-turkey approaches to smoking cessation
LDL and HDL
Saturated fat, unsaturated fat, and trans-fatty acids
Young children's and high schoolers' dietary behaviors
Original food pyramid and redesigned food pyramid
Yo-yo dieting and healthy dietary intake
Body fat and body weight
Endurance, strength-building, and stretching
Genital herpes, syphilis, and HIV
Abstinence only and abstinence plus programs

UNIT B INSTRUCTOR-NOTES CONCEPTUAL PAIRS

Piaget's early training and professional contributions
Assimilation and accommodation
Object permanence and conservation
Sensory areas and association areas
Synapses and neurons
Prenatal and postnatal brain development
Longevity and cost-effectiveness of preschool programs
Intelligence testing and curriculum-based assessment
Divergent and convergent thinking
Disposition and ability to think critically
HPLCT, LPLCT, and HPHCT
Information bases of high and low critical thinkers
Learning style and instructional approaches related to learning disabilities
Whole language and phonics
Phonemic awareness and alphabetic coding
Task analysis and immersion
Whole language and direct instruction

UNIT C INSTRUCTOR-NOTES CONCEPTUAL PAIRS

Nomothetic and hierarchical models of self-concept
Taxonomic and compensatory models of self-concept
BFLPE and LFBPE
Self-esteem as a precondition and product of learning
Locus of control and self-efficacy
Humanism and behaviorism
Feelings and behaviors
Unidirectional and reciprocal reinforcement
Intrinsic and extrinsic reinforcement
Participation contingency and quality contingency
Deficiency and growth needs
Stress and stressor
Problem-focused and emotion-focused coping
Functional and dysfunctional thinking
Tokens and backup rewards
Response cost and differential reinforcement
Biofeedback and Ritalin
Theta, beta, and sensory-motor waves
Suicide rates in adolescence/early adulthood and old age
Suicidal patterns among males and females

UNIT D INSTRUCTOR-NOTES CONCEPTUAL PAIRS

Slavin and Kohn
STAD, Jigsaw II, and TGT
Controlled floundering and direct instruction
Competition and cooperation
Conventional academic work and creative problem-solving
Group recognition and individual accountability
Academic and social effects of cooperative learning
Effects of cooperative learning on high, average, and low achievers
Benefits of cooperative learning for collaborative and individualistic students
Task structure and reward structure
Basic, more advanced, and most advanced cooperative skills
Academic and social benefits of peer tutoring
Boys' and girls' patterns of interpersonal relationships
Self-report and projective techniques
Sociometrics, behavior rating scales, and classroom observation
Structured controversy and free controversy
Authoritarian and authoritative parenting
Indulgent and uninvolved parenting

UNIT E INSTRUCTOR-NOTES CONCEPTUAL PAIRS

Moral conduct and moral reasoning
Circles test and Defining Issues Test
Clinical interview and Defining Issues Test
Laboratory assessment and applied behavior assessment
Beliefs about cheating and self-reports of cheating
Moral reasoning and cheating
Volunteerism and service learning
Environmentalism and religious beliefs
U.S. consumption of resources and voluntary simplicity lifestyle
Gun control laws and gun violence
Perceived U.S. support for Israel and Islamic cultures
Religiosity and support for human rights
Nationalism and patriotism
Ethnic pride and ethnic hostility
Christian fundamentalism and militarism

VITA

Haley Crisp Turner was born in Knoxville, TN on June 1, 1980. She was raised in North Knoxville. She graduated from Halls High School in 1998. From there, she went to East Tennessee State University in Johnson City, TN where she pursued a degree in Psychology. She received a B.S. in Psychology in 2002. Haley is currently pursuing her doctorate in education with a concentration in school psychology at the University of Tennessee, Knoxville, TN.